DOCUMENT RESUME

ED 091 412 TM 003 624

AUTHOR Katzenmeyer, Conrad G.; And Others
TITLE A Model for the Development and Evaluation of Placement Tests for Objective Based Curriculum Management Systems.
PUB DATE Apr 74
NOTE 13p.; Paper presented at American Educational Research Association Annual Meeting (Chicago, Illinois, April 15-19, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Diagnostic Tests; Educational Diagnosis; Elementary School Students; Grouping (Instructional Purposes); Individualized Instruction; *Models; *Placement; *Test Construction; *Tests
IDENTIFIERS WDRSD; Wisconsin Design for Reading Skill Development

ABSTRACT

In many objective-based curriculum management systems, students' curricular activities are carefully directed by their own performance through extensive pretesting. When implementing such programs, however, there are often only rough criteria for appropriate leveling of students, necessitating extensive retesting. This paper outlines a model for the development and evaluation of a placement test for the Word attack area of the Wisconsin Design for Reading Skill Development. A thirty-item placement test was constructed and tried out in two elementary schools prior to program implementation. Development strategies and effectiveness of the placement test in minimizing leveling errors are discussed. (Author)

A MODEL FOR THE DEVELOPMENT AND EVALUATION OF PLACEMENT TESTS FOR OBJECTIVE BASED CURRICULUM MANAGEMENT SYSTEMS

Conrad G. Katzenmeyer, Ph.D.
R & D Center for Cognitive Learning
University of Wisconsin

Deborah M. Stewart
R & D Center for Cognitive Learning
University of Wisconsin

Mary R. Quilling
University of Massachusetts

In many objective-based curriculum management systems, students' curricular activities are carefully directed by their own performance through extensive pretesting. When implementing such programs, however, there are often only rough criteria for appropriate leveling of students, necessitating extensive retesting. This paper outlines a model for the development and evaluation of a placement test for the Word Attack area of the Wisconsin Design for Reading Skill Development. A thirty-item placement test was constructed and tried out in two elementary schools prior to program implementation. Development strategies and effectiveness of the placement test in minimizing leveling errors are discussed.

Paper presented at AERA Meeting, Chicago, April, 1974.

Objectives

Objectives-based curricula and curriculum management systems are becoming more common as schools move toward individualization and competency-based education. In curriculum management systems of this nature, such as the Wisconsin Design for Reading Skill Development (Otto and Askov, 1972), a student's curricular activities are carefully directed by his own performance, usually through an extensive pretesting procedure. Yet when such a management is first implemented in a school, there are often only age-grade and teacher judgment criteria to guide initial placement, which can lead to a large retesting rate which results in a considerable loss of time and resources.

The purpose of this paper is to discuss and evaluate a preliminary model for the development of a curriculum placement test that will aid in initial leveling of students prior to the pretesting or "break-in" testing at any level. A successful placement test should be able to substantially lower the retesting rate indigenous to the system.

Theoretical Framework

The unique and critical role of placement tests is recognized in models of educational assessment (Hillson and Bongo, 1971). They are a type of diagnostic test used for determining the degree of mastery of program objectives already attained (Bloom, Hastings, and Madaus, 1971). A placement test must accurately reflect the program's objectives, yet must cover a wider range of those objectives than any specific level of the program would contain. As such they represent the first stage of a two stage sequential testing strategy (Cronbach and Gleser, 1965).

A particular problem for many placement tests occurs when the base rate of correct placement without the test exceeds 50%. As discussed by Meehl and Rosen (1955), placement tests must be extremely valid to be useful when the behavior to be predicted is already reasonably well

predicted without the test. In this study, correct placement without the test is approximately 75%.

Methodology

To be successful, a placement test must provide a scoring system that will minimize the number of students requiring retesting at a higher or lower level, but must be brief and easily administerable. The first step was to construct a short test representing diverse elements of the 40 battery tests of the WDRSD Word Attack area. As the Word Attack battery ranges in difficulty from Level A (Prereading level) to Level D (completion of word attack skills) no student takes the full battery of 40 tests in any single year. Therefore, it was necessary to evaluate tests at a given level and combine by extrapolation to a scaling model. Tests at each level were examined to see which were the most successful in predicting the decision of correct or incorrect placement. Then items within each scale were evaluated in regard to the same decision. In addition, scales were included that reflect a representative sample of phonic and structural strands within Word Attack. The final placement test consisted of 30 items - 5 items from a Level A test, 10 from two Level B tests, 10 from two Level C tests, and 5 from a Level D test. The means of these tests formed a scaling pattern, and items were chosen to retain this scaling pattern with the shorter subtests.

Data Source

In the Spring of 1973 the Word Attack Placement Test was administered to all students in two elementary schools (a suburban school in Wisconsin and an urban school in California), prior to break-in testing for the Wisconsin Design. These schools were not allowed to use the placement

test information for initial leveling. Students were scored for mastery (80%) as well as total score on each subscale.

Results

Placement Test

Means, standard deviations and reliabilities for the six subscales are given in Table 1. Separate results are listed for grades K-4 and grades 1-4, as Kindergarteners were not required to take the final three subscales unless they could read. The Kindergarteners' scores were retained for pattern analysis, however, thus inflating reliability estimates.

Pattern analysis according to mastery scores of the six subscales is given in Tables 2 and 3. For the total sample, 639 of 776 or 82.34% of the students conformed to the scale pattern expected. Of the non-scale patterns found (see Table 3) 10 of the Kindergarteners failed to master subscale 1 but did mastery subscale 2 (pattern #1) and 55 of the students mastered all subscales but Subscale 5 (pattern #20). Based on this information and the fact that neither Subscale 1 or Subscale 6 discriminated well in the range of students available for this study, both subscales were dropped from further analysis. This eliminated non-scale patterns 1, 3, 18 and 20, and left 716 of 776 or 92.26% of the students conforming to the scale patterns.

Relationship to Break-in Testing.

Results of the break-in tests for the full battery, using standard, non-placement test guidelines, are given in Table 4. A student is considered inappropriately leveled if he masters 0 or only 1 scale at a level (test down) or masters all or all but 1 test at a level (test up). The overall error for initial placement of 26.6% was very close to the expected 25%, with

somewhat larger error rates at the lower battery levels. One difficulty that surfaced at this time in regard to this sample was that almost all inappropriate placements were"test ups." This was undoubtedly due to the fact that the break-in testing occurred late in the school year; a different pattern of errors would be expected if break-in had occurred in the fall. Therefore, the results obtained in this study apply only to Spring implementation of Word Attack and the study will need to be repeated next fall.

Comparisons of the placement test results with the full-battery results are given in Table 5. Conditional probabilities were computed separately for the appropriately and inappropriately placed students at each test level. Predictions from the Placement Test were run for each subscale, combinations of subscales, and for test score totals in addition to combinations. In order to simplify the table, only subscales and subscale combinations are listed.

Results indicate that the Placement Test could markedly improve placement at Level A, using mastery of subscale 2 only. Level AB was totally unpredictable; the Placement Test Subscale 3 had negative discrimination at this level. Detailed analysis of the sample suggested that the problem arose in two classes of Kindergarteners where a number of students could master enough tests at A-B to be judged test-ups, but had not had the material in Subscale 3 and thus could not master it. Further work will need to be done at A-B in order to make the Placement Test useful.

Predictions at Levels B and C were somewhat above the base rates, when total test scores were included with subscale mastery scores. The results were not particularly striking, however, and it seems likely that at these levels predictions will have to be made in one direction only.

The Placement Test will probably be recommended as a threshold variable; if a student does not achieve mastery of a number of subscales and/or a total score of a certain level, a prediction can be made that the student should not be tested up. However, a score at or above the levels set will have to be interpreted as a sign to consider further information before testing up.

Due to the base rate of appropriate placement, Level D was not predictable from the Placement Test. This is not a serious problem as the decision to test up from Level D means that the child has completed all Word Attack skills, and thus is the type of decision which should be made from the total battery, not from a Placement Test.

Implications

Based on the data collected, the model for developing placement tests for objectives-based curriculum management systems presented here has been reasonably successful. Predictions for three of the five curriculum levels were better than the base rate, although the results obtained in this study need cross-validation and further investigation at another time in the school year. Further, the Placement Test was developed with no additional test construction and proved to be a good instrument, both in terms of internal characteristics and in scale patterns. It may well be that the scale distances of the subscales need to be adjusted, as there was more than the expected mastery overlap for the two most difficult scales, yet the Placement Test may well need somewhat greater "top" than was available with only four subscales.

Possibly the most important finding in this study was that the Placement Test could provide highly accurate information in only one direction. Given the base rate of approximately 75%, the Placement Test was quite effective

Table 2
Number of Students
Conforming to Scale Patterns
(0=nonmastery 1=mastery)

Scale Patterns	Number of Students	Percent of Students
000000	22	2.84
100000	59	7.61
110000	93	11.98
111000	106	13.65
111100	145	18.68
111110	84	10.83
111111	130	16.75
Total	639	82.34

Table 1
Means, Standard Deviations and
Reliabilities of the Subscales of
the Word Attack Placement Test

Subscale	Mean		Standard Deviation		Reliability	
	K-4	1-4	K-4	1-4	K-4	1-4
A	4.82	4.94	.60	.32	.64	.55
B	4.45	4.72	1.15	.72	.79	.63
C	3.82	4.35	1.69	1.16	.86	.73
D	3.20	3.79	2.04	1.63	.91	.83
E	2.27	2.69	1.82	1.67	.79	.71
F	2.06	2.44	1.79	1.69	.78	.71
Total	20.61	22.95	7.40	5.36	.94	.88

Table 3

Number of Students by Grade Having Nonscale Patterns for the Word Attack Placement Test

Nonscale Patterns	K	1	2	3	4	Total
1. 010000	10	2	1			13
2. 010100				1		1
3. 011100		1				1
4. 000101				1		1
5. 100100		1		1		2
6. 101000		1	4			5
7. 100111			1			1
8. 101010			1			1
9. 101100			3			3
10. 101110		1	2			3
11. 101101					1	1
12. 101111					1	1
13. 110010		1				1
14. 110100		6	1	3	1	11
15. 110110			4	1	2	7
16. 110101				5	2	7
17. 110111				3	3	6
18. 111001		2	4	2		8
19. 111010		3	5	1		9
20. 111101		3	15	34	3	55
	10	21	41	52	13	137
Total # Taking Test	122	221	186	198	49	776
% of Nonscale Patterns	8.2	9.5	22.0	26.3	26.5	17.6

Table 4
Results of Full Battery
Break-in Testing

	Appropriately Placed	Inappropriately Placed	% Appropriate
A	58	41	58.6
AB	12	16	42.9
B	107	38	73.8
C	154	56	73.3
D	99	5	95.2
Total	430	156	73.4

Table 5

Conditional Probabilities of Proper Placement
Appropriately and Inappropriately
Placed Students Using Several Soring Techniques
for the Word Attack Placement Test

Breakpoint Test Results (n in parentheses)	Placement Test Subscales							Overall Best Prediction
	SS2	SS3	SS4	SS5	SS2&3	SS2,3&4	SS2,3,4,&5	
A Appropriate (58)	.73	1.00	1.00	1.00	1.00	1.00	1.00	.72
Inappropriate (41)	.71	.02	.00	.00	.02	.00	.00	
AB Appropriate (12)	.17	.33	.67	1.00	.42	.67	1.00	.39
Inappropriate (16)	1.00	.38	.00	.00	.38	.00	.00	
B Appropriate (107)	.02	.19	.43	.92	.20	.67	.95	.81
Inappropriate (38)	.97	1.00	.89	.42	.97	.87	.37	
C Appropriate (154)	.04	.10	.20	.62	.12	.25	.66	.79
Inappropriate (56)	1.00	1.00	1.00	.80	1.00	1.00	.80	
D Appropriate (99)	.06	.09	.12	.34	.14	.22	.43	.46
Inappropriate (5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

in providing a threshold level below which it could be said with considerable certainty that the student was properly leveled, but the decision to "test up" when the threshold score was exceeded could not be made with similar accuracy. It remains to be seen whether this option of the placement test as a threshold measure will be supported in a new sample that contains a larger portion of "test downs".

References

Bloom, B., Hastings, T., and Madaus, G. Handbook of Formative and Summative Evaluation of Student Learning. Chicago: McGraw-Hill, 1971.

Cronbach, L. J. and Gleser, G. Psychological Tests and Personnel Decisions. Urbana: University of Illinois Press, 1965.

Hillson, M. and Bongo, J. Continuous-Progress Education. Palo Alto: Science Research Associates, 1971.

Meehl, P. E. and Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psychological Bulletin: 1955, 52, 194-216.

Otto, W. and Askov, E. The Wisconsin Design for Reading Skill Development: Rationale and Guidelines. Minneapolis: National Computer Systems, 1972.